LA-UR-22-20459

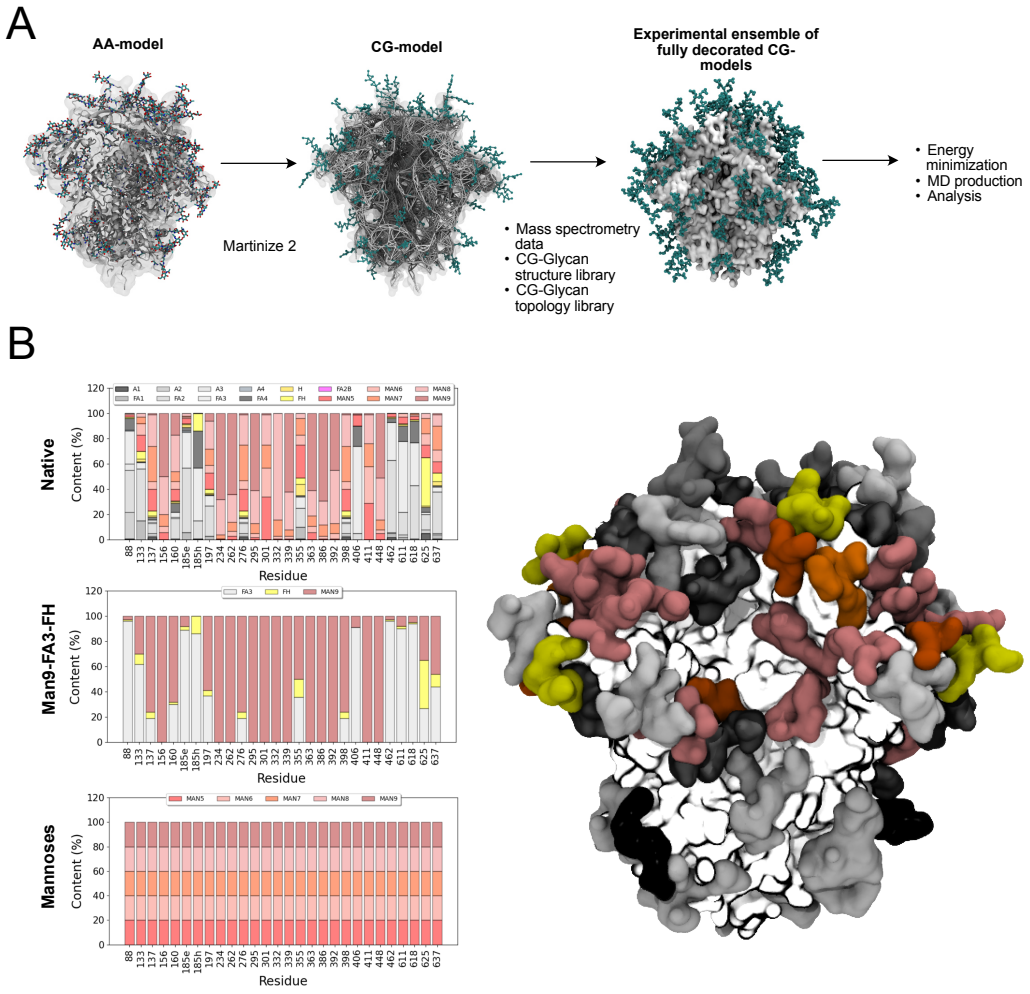
Approved for public release; distribution is unlimited.

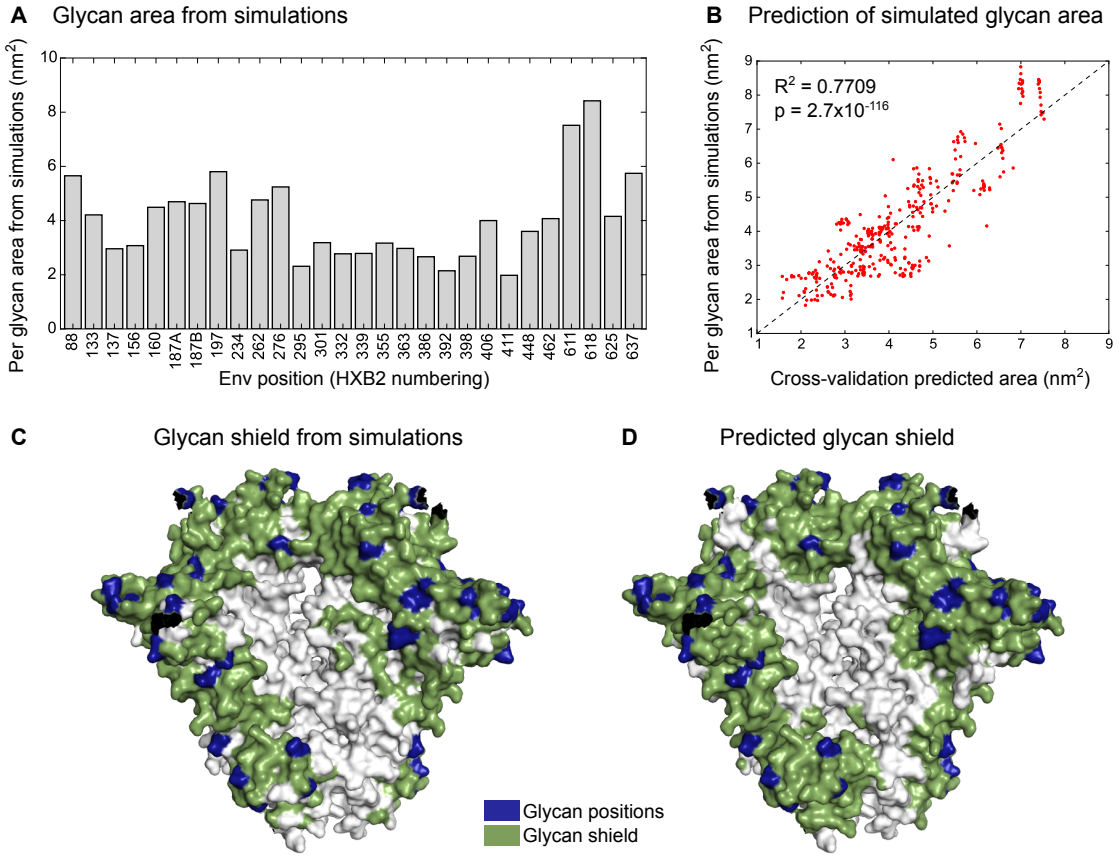
Title:	Large scale MD to predict Epitope regions in HIV Env
Author(s):	Lopez Bautista, Cesar Augusto
Intended for:	Report
Issued:	2022-01-20

Large scale MD to predict Epitope regions in HIV Env

In silico generation of glycosylation biophysical information

AI, deep learning and glycan “Shield” prediction





Abstract

Highly dense carbohydrates located on the surface of the HIV Env protein play a key role in immune evasion. Such evolutionary adaptation hampers any attempt to obtain a full mechanistic understanding of the role played by the glycans in protecting the virus against an effective immune response. Moreover, and due to their chemical variability, an accurate molecular understanding of the so called “glycan shield” is still limited by the lack of effective resolution of state-of-the-art experimental technics. Here, we have used extensive computational modelling in order to fill this gap, addressing the presence of a large glycan variability as observed experimentally. Based on an automated pipeline, we were able to assemble, set-up and simulate *via* Molecular dynamics hundreds of different glycosylated Env variants at nearly atomic resolution, recapitulating the glycosylation distributions observed experimentally.

Results from these simulations were subjected to machine learning and very accurate prediction of simulation derived glycan shielding areas of each glycan as a function of static sequence features. Such predictive models of per-glycan shielding, incorporating both glycan dynamics and heterogeneity, were used to develop a novel sequence-based glycan shield mapping strategy. Parallel to these studies, we also developed an accurate machine learning approach to predict glycan heterogeneity data using sequence features and found good prediction accuracy.